



## Integration of Known DNA, RNA and Protein Biomarkers Provides Prediction of Anti-TNF Response in Rheumatoid Arthritis

results from the COMBINE study

**Folkersen, Lasse; Brynedal, Boel; Marcela Diaz-Gallo, Lina; Ramsköld, Daniel; Shchetynsky, Klementy; Westerlind, Helga; Sundström, Yvonne; Schepis, Danika; Hensvold, Aase; Vivar, Nancy**

*Total number of authors:*  
19

*Published in:*  
Molecular Medicine

*Link to article, DOI:*  
[10.2119/molmed.2016.00078](https://doi.org/10.2119/molmed.2016.00078)

*Publication date:*  
2016

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

### *Citation (APA):*

Folkersen, L., Brynedal, B., Marcela Diaz-Gallo, L., Ramsköld, D., Shchetynsky, K., Westerlind, H., Sundström, Y., Schepis, D., Hensvold, A., Vivar, N., Eloranta, M-L., Rönnblom, L., Brunak, S., Malmström, V., Catrina, A., G. W. Moerch, U., Klareskog, L., Padyukov, L., & Berg, L. (2016). Integration of Known DNA, RNA and Protein Biomarkers Provides Prediction of Anti-TNF Response in Rheumatoid Arthritis: results from the COMBINE study. *Molecular Medicine*, 22, 322-328. <https://doi.org/10.2119/molmed.2016.00078>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Integration of Known DNA, RNA and Protein Biomarkers Provides Prediction of Anti-TNF Response in Rheumatoid Arthritis: Results from the COMBINE Study

Lasse Folkersen,<sup>1,2</sup> Boel Brynedal,<sup>3</sup> Lina Marcela Diaz-Gallo,<sup>2</sup> Daniel Ramsköld,<sup>2</sup> Klementy Shchetynsky,<sup>2</sup> Helga Westerlind,<sup>3</sup> Yvonne Sundström,<sup>2</sup> Danika Schepis,<sup>2</sup> Aase Hensvold,<sup>2</sup> Nancy Vivar,<sup>2</sup> Maija-Leena Eloranta,<sup>4</sup> Lars Rönnblom,<sup>4</sup> Søren Brunak,<sup>1</sup> Vivianne Malmström,<sup>2</sup> Anca Catrina,<sup>2</sup> Ulrik GW Moerch,<sup>5\*</sup> Lars Klareskog,<sup>2\*</sup> Leonid Padyukov,<sup>2\*</sup> and Louise Berg<sup>2\*</sup>

<sup>1</sup>Department of Bioinformatics, Technical University of Denmark, Lyngby, Denmark; <sup>2</sup>Unit of Rheumatology, Department of Medicine Solna, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden; <sup>3</sup>Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; <sup>4</sup>Department of Medical Sciences, Uppsala Universitet, Uppsala, Sweden; and <sup>5</sup>Current affiliation: AbbVie, Copenhagen, Denmark; \*These senior authors contributed equally

**OBJECTIVE:** In rheumatoid arthritis (RA) several recent efforts have sought to discover means of predicting which patients would benefit from treatment. However, results have been discrepant with few successful replications. Our objective was to build a biobank with DNA, RNA and protein measurements to test the claim that the current state-of-the-art precision medicine will benefit RA patients. **METHODS:** We collected 451 blood samples from 61 healthy individuals and 185 RA patients initiating treatment, before treatment initiation and at a 3 month follow-up time. All samples were subjected to high-throughput RNA sequencing, DNA genotyping, extensive proteomics and flow cytometry measurements, as well as comprehensive clinical phenotyping. Literature review identified 2 proteins, 52 single-nucleotide polymorphisms (SNPs) and 72 gene-expression biomarkers that had previously been proposed as predictors of Tumor Necrosis Factor (TNF) inhibitor response ( $\Delta$ DAS28-CRP). **RESULTS:** From these published TNFi biomarkers we found that 2 protein, 2 SNP and 8 mRNA biomarkers could be replicated in the 59 TNF initiating patients. Combining these replicated biomarkers into a single signature we found that we could explain 51% of the variation in  $\Delta$ DAS28-CRP. This corresponds to a sensitivity of 0.73 and specificity of 0.78 for the prediction of three month  $\Delta$ DAS28-CRP better than  $-1.2$ . **CONCLUSIONS:** The COMBINE biobank is currently the largest collection of multi-omics data from RA patients with high potential for discovery and replication. Taking advantage of this we surveyed the current state-of-the-art of drug-response stratification in RA, and identified a small set of previously published biomarkers available in peripheral blood which predicts clinical response to TNF blockade in this independent cohort.

**Online address:** <http://www.molmed.org>

**doi:** 10.2119/molmed.2016.00078

## INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory disease resulting primarily in chronic inflammation and destruction of symmetric joints. Persons

with RA are often in working age and the accompanying fatigue significantly affects working capacity. Ongoing joint destruction is however seen in more than a third of patients after initiation

of a treatment regime. There is currently no generally accepted way to predict treatment efficacy in individual patients, so medications are prescribed according to consensus recommendations. First-line treatment is typically methotrexate (MTX), an inhibitor of protein and nucleic acid synthesis that leads to inhibition of immune cells (1). Around 30% of RA patients do not respond to MTX and are then prescribed a combination of MTX and an anti-TNF blocking agent. TNF drives the inflammation within the joint, and blocking reduces immune cell infiltration and immune mediated joint destruction (2). About 30% of patients prescribed their first anti-TNF therapy

\*LK and LB contributed equally to this work.

**Address correspondence to** Lasse Folkersen. Center for Biological Sequence Analysis, Department of Bioinformatics, Technical University of Denmark, Building 208, 2800 Lyngby, Denmark. Phone: +45-29936063; Fax: +45-45931585; E-mail: lasfol@cbs.dtu.dk. Submitted March 23, 2016; Accepted for publication June 24, 2016; Published Online (www.molmed.org) August 15, 2016.

fail to respond, upon which other biologic therapies are prescribed. The current challenge of translational research in this area is to better utilize the treatment options that already exist, in a personalized or stratified manner. Several groups have attempted to use transcriptomics (3–10), genetics (11–13) and proteomics (14), as well as better use of clinical data (15) to predict treatment response, particularly for TNF blockade. Success has been limited with virtually no findings validated in independent material, and no biomarker for prediction of response is currently used in clinical practice (11). In any such study the collection of relevant biological samples is of key importance, and independent validation of results is necessary for further research.

We therefore set out to compile the COMBINE biobank of samples from RA patients that included global profiling of transcriptomics, genetics, proteomics, flow cytometry and clinical information. With this unique resource, as a first stage we performed a complete quantification of all previously suggested anti-TNF response biomarkers: to investigate how well precision medicine would actually work, given the input of all previous knowledge on RA precision medicine that we have today. To our knowledge this biobank is currently the largest collection of such multi-omics data from RA patients. We present this as an essential guidance in the highly discrepant field of drug response stratification research, as a resource for combining the findings of the many excellent studies already published.

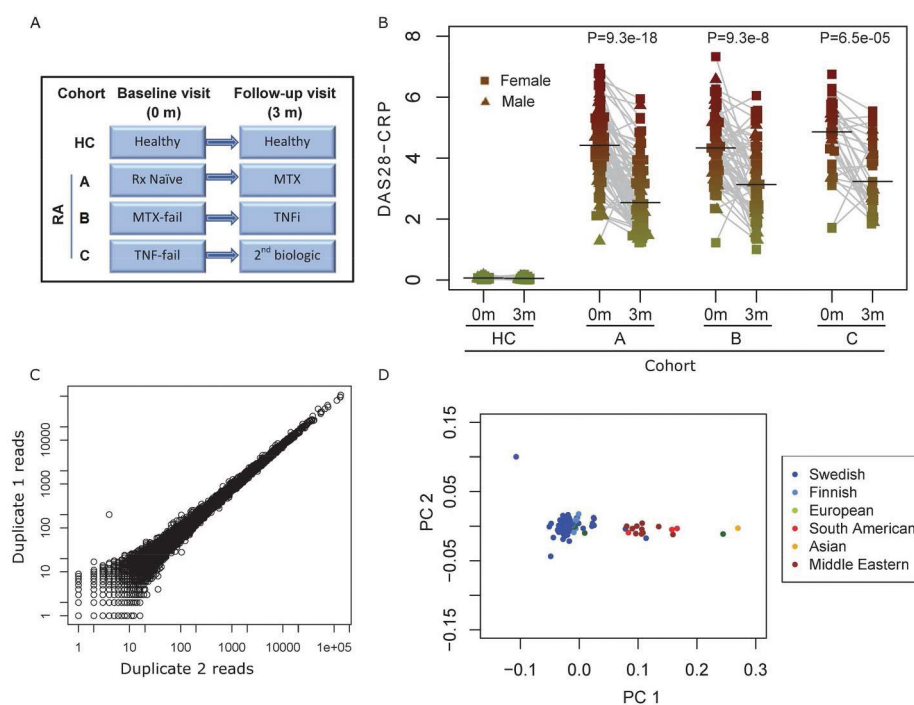
## MATERIALS AND METHODS

### Study Design and Sample Collection

The COMBINE biobank was generated after written informed consent from all participants had been obtained according to the declaration of Helsinki and with approval by the Stockholm (number 2010-351-31-2) and Uppsala (2009-013) Regional Ethics Committees. The key inclusion criteria were patients with rheumatoid arthritis, according to

the ACR 1987 or the 2010 ACR/EULAR criteria, who were undergoing change or start of a new treatment regimen at the Rheumatology Clinic, Karolinska University Hospital, Stockholm from February 2011 to May 2013. Our cohort includes 3 patients groups (Figure 1A): one group of patients with symptoms initiating no more than 14 months ago and initiating MTX treatment (*cohort A*), one group of early RA patients initiating anti-TNF therapy of any type (*cohort B*) and one group of patients initiating a second-line biologics treatment, which could be either anti-TNF or other biologics (*cohort C*). Additionally, healthy controls were recruited from the Swedish Blood Centre service in Uppsala (*cohort HC*), with age and

sex as closely matched with patient groups as possible. The blood samples were collected at the baseline visit (*timepoint 0m*) and at a follow-up visit approximately 3 months later (*timepoint 3 months, 95.3 d, standard deviation (SD) 10.5 in patient groups, but larger variation in healthy controls*). At both visits patients underwent a full clinical evaluation, and data were recorded on number of swollen and tender joints, general health, C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR). From these the disease activity score (DAS28) was calculated, and we have specifically used the DAS28 including CRP levels in our analysis. At the clinical visits information regarding smoking status, all



**Figure 1.** A) Overview of study design and cohort setup. B) DAS28-CRP score stratified by cohort and visit time, with an assumed value of 0 for the HC cohort. Each dot shows one sample. The Y-axis shows DAS28-CRP level and the X-axis shows cohort membership and time point. Grey lines connect paired time-points, and horizontal black bars indicate median value. Color-coding corresponds to figure 2C. C) Reproducibility of RNA-seq expression levels from technical replicates, in different library preparations with three months separation. Each dot shows the read count of one single gene. The Pearson correlation coefficient of log transformed counts was 0.991 ( $P < 2e-16$ ). D) Principal components analysis of genome-wide genotype profile and self-reported country of origin of individuals. Each dot shows one patient, color-coded by self-reported country of origin. TNFi: TNF inhibitor, MTX: Methotrexate, Rx naïve: no methotrexate initiation.

current and previous medications, cell counts of all subsets of leukocytes, as well as samples for RNA sequencing of peripheral blood mononuclear cells (PBMC), flow cytometry analysis of PBMC and whole blood and serum and plasma analyses for protein biomarkers were taken. At baseline, patients were additionally evaluated for anticitrullinated protein antibody (ACPA) and rheumatoid factor (RF) positivity, and gave DNA samples for microarray genotyping and human leukocyte antigen (HLA) haplotype characterization. A total of 492 blood samples from 246 different individuals were collected.

### Serum Protein Measurements

Proteins were measured using multiple commercial protein biomarker panels. These included 62 analytes in the HumanMAP assay (Myriad RBM), 12 analytes in the VectraDA (Crescendo) as well as 33 autoantibodies using a fluorescence enzyme immunoassay technique from Phadia (Thermo Scientific). These plasma protein analyses were performed at external centers according to the standard protocols of the labs as indicated. The proteins BAFF and CXCL13 were analyzed using RnD Systems and APRIL using AH Diagnostics ELISA kits. Depending on assay type, between 422 and 451 samples were successfully analyzed.

### RNA Purification and Sequencing

RNA was purified from PBMC samples collected using CPT tubes (BD Biosciences) and purified using isopropanol extraction. The quality of each RNA sample was characterized using an Agilent Bioanalyser (2100 total RNA pico chip). RNA was sequenced at the Aros Applied Biotechnology center, using an Illumina HiSeq 2000, the TruSeq RNA sample preparation kit and 2 × 100 base pair (bp) paired end setup. Samples were randomized to flow cells with focus on obtaining balanced distribution of cohort type and drug response magnitude. Mean read depth was 15.7 million read-pairs per sample. Sample quality was inspected using read depth, fastqc quality calls

and a check for Y-chromosome expression. Altogether 17 RNA samples were re-sequenced due to quality concerns and finally 2 samples were omitted from the study. A total of 360 samples were successfully RNA sequenced. One of the samples was divided in two aliquots before library preparation to investigate reproducibility of RNA-sequence expression.

A pre-filtering on quality of reads using the *fastqc* package (0.10.1) was applied, removing adaptors and applying *fastqc\_quality\_trimmer* (-q 30, -p 85) and *fastqc\_quality\_filter* (-t 30 -L 40). This was followed by *tophat2* (2.0.10) alignment to the ensemble GRCh37 genome together with the -gtf switch with the hg19 UCSC transcripts (16). Gene expression was quantified using *htseq* followed by TMM-normalization, mean-scaling and log2 transformation as implemented in the *edgeR* package (3.8.6) (17) as recommended by Dillies *et al.* (18).

### DNA Purification and Microarray Genotyping

Blood was collected in Vacutainer citrate tubes (BD) and DNA purified according to the QIAasympyphony 400-ul buffy coat protocol. The DNA was hybridized using the HumanOmniExpress BeadChip Kit and Illumina OmniExpress arrays (12v1). Y-chromosome call rates were compared with recorded sex and found to be mismatched for three samples which were excluded. Because of these sex concordance based sample switches, coding genotypes in highly expressed genes were compared with those obtained from RNA-sequence data using the *allelicImbalance* package (1.4.2) (19). This resulted in an additional 3 omitted samples due to mismatch of genotypes call. Altogether 230 individuals were successfully genotyped. Principal components analysis was performed using the *made4* package (1.40.0) on all measured genotypes. The first 10 principal components were investigated – the eighth and ninth showed the best resolution of larger ethnicity groups, the remaining

components instead being driven by smaller outliers or sex.

Imputation of non-measured SNPs was performed using the Mach1.0 imputation algorithm (1.0.18.c), with separate phasing in minimac (2012-11-16). For reference the 1,000 genomes project EUR was used (release\_v3.20101123). This provided 7.7 million SNPs available for analysis at Rsq quality values above 0.7.

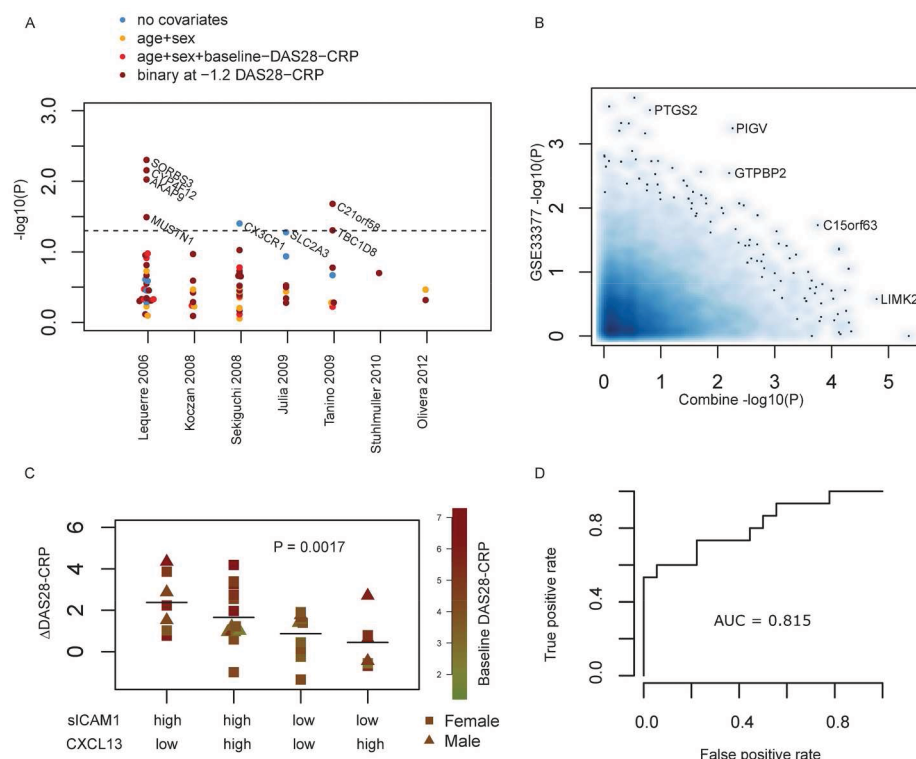
### Association Between Biomarkers and DAS28-CRP

The primary outcome variable was defined as difference in DAS28-CRP three months after initiation of treatment with TNF inhibitors. To account for variance in methodology of the included discovery studies, the association between biomarker level at baseline and  $\Delta$ DAS28-CRP was calculated using 4 different models: one model which used a Student *t* test between patients with  $\Delta$ DAS28-CRP values above and below -1.2, and three linear regression models, one without covariates, one with age and sex as covariate and one with age, sex and baseline DAS28-CRP value as covariate. One exception was the protein data in Figure 2C, which was handled using the same split by median values described in Dennis *et al.* (14). This was followed by a linear regression with the indicated four protein level categories as predictor variable.

The multiple testing burden of the study corresponds to the count of literature identified biomarker candidates. Details of the literature search are provided in Supplement 1. A total of 2 + 52 + 72 biomarkers were included in the restricted search space. For these, *P* values are always reported as raw uncorrected values. The comparison plot in Figure 2B was made using the smoothScatter from the R-graphics base-package and the receiver operated characteristics plot in Figure 2D was calculated using a ROCR package. All calculations were performed in R version 3.1.2.

*All supplementary materials are available online at [www.molmed.org](http://www.molmed.org).*





**Figure 2.** A) Replication of mRNA as stratification biomarkers for anti-TNF-response in 59 patients from cohort B. Each dot shows one gene, and the Y-axis shows association strength. For each of 7 published studies, all reported genes were investigated for association to anti-TNF-response. Four different models were calculated: three linear regressions with different covariate setups, and one binary-response/no-response setup set at  $\Delta$ DAS28-CRP  $< -1.2$ . The strongest covariates reported in the respective studies are indicated by the color code in the legend. B) Comparison with mRNA expression data from Toonen *et al.* (GSE33377). The X- and Y-axis shows the  $\log_{10}(P)$ -value for association with TNFi response for each measured gene, from the Combine and the GSE33377 data, respectively. The compared  $P$ -values were calculated based on binary response/no-response. C) Replication of serum protein expression biomarkers sICAM1 and CXCL13 previously reported by Dennis *et al.* (14). Protein levels were divided by median values for each of the two proteins as indicated along X-axis. The  $P$ -value was calculated for a linear regression over the four levels. Each dot shows one patient. Color coding and symbol shape indicate baseline DAS28-CRP and gender. D) ROC plot illustrating the ability to estimate if a patient will have more or less than  $-1.2$   $\Delta$ DAS28-CRP values, based on the fitted values of a model containing the 11 predictive biomarkers. This curve shows how different threshold levels will affect the rate of true and false positives predicted: the interpretation of the area under curve (AUC) is that 81.5% of randomly selected TNFi responders will have a higher test result than randomly selected non-responders.

## RESULTS

In the COMBINE biobank we collected 492 blood samples from 246 individuals, and applied a range of high-throughput measurement techniques to most of them (Table 1). The study design is illustrated in Figure 1A, and the DAS28-CRP changes in Figure 1B.

Herein is also a clear illustration of the main issue of non-response to common drug types: only 47% of the patients had  $\Delta$ DAS28-CRP scores better than  $-1.2$  for evaluation at 3 months after treatment initiation. Figure 1C shows the precision of the RNA-sequence and Figure 1D illustrates the self-reported and genotype

derived ethnicity of participants, which could be an important aspect of the likelihood of reproducing SNP-based drug stratification markers.

## Transcriptomics-Based Anti-TNF-Response Association

Using keywords anti-TNF, RA and gene expression profiles in blood cells we identified eight other studies containing similarly designed setups of gene expression profiles and data for anti-TNF-response transcriptomics (3–10). For each of these studies we sought to replicate the findings in the COMBINE biobank (Figure 2A). From a total of 72 candidates of previously proposed blood gene expression biomarkers, we observed that the genes *SORBS3*, *AKAP9*, *CYP4F12*, *MUSTN* (4), *CX3CR1* (5), *SLC2A3* (6), *C21orf58* and *TBC1D8* (7) could be replicated as stratification biomarkers at  $P < 0.05$  from the COMBINE dataset, and with consistent direction. One study had provided genome-wide expression data from whole blood, accession GSE33377 (10), and we additionally attempted to quantify how much overlap of findings there was when considering all genes (Figure 2B). Ideally one would expect several genes to have strong significance in both studies, which was not the case. There is thus a limited ability to reproduce findings between cohorts of otherwise similar size, which emphasizes the need for more validation studies in the field and argues for integration of data obtained from several patient cohorts.

## Proteomics-Based Anti-TNF-Response Association

Next, we performed a similar analysis for recent findings that sICAM1 and CXCL13 serum protein levels could be used for stratification of anti-TNF-response (14). When dividing the anti-TNF initiating patients (cohort B) into medians of sICAM1 and CXCL13 levels and comparing the results of the four groups as done by Dennis *et al.*, we found a strong replication of the original findings (Figure 2C). Interestingly, correlation and other approaches that were

**Table 1.** Demographic information for patients in the biobank.

	HC-0m (n = 61)	HC-3m (n = 38)	A-0m (n = 92)	A-3m (n = 81)	B-0m (n = 59)	B-3m (n = 57)	C-0m (n = 34)	C-3m (n = 30)
Age (y)	47 ± 13		60 ± 15		55 ± 13		55 ± 15	
Female	61% (37)		72% (66)		69% (41)		71% (24)	
CPP-positive			60% (55)		78% (46)		79% (27)	
RF-positive			54% (50)		66% (39)		76% (26)	
DAS28-CRP			4.5 ± 1.2	2.8 ± 1.2	4.4 ± 1.2	3.2 ± 1.2	4.7 ± 1.2	3.4 ± 1.2
CRP (mg/mL)			14 ± 24	5.7 ± 15	11 ± 22	5.9 ± 14	12 ± 15	10 ± 14
ESR (mm)			32 ± 22	19 ± 14	20 ± 13	17 ± 13	28 ± 19	24 ± 15
Pain (VAS mm)			52 ± 27	22 ± 22	45 ± 26	30 ± 28	50 ± 22	42 ± 28
CD3 + T cells of PBMC (%)	60 ± 7.3	59 ± 7.5	51 ± 18	53 ± 14	49 ± 19	55 ± 13	55 ± 17	61 ± 8.4
CD19 + B cells of PBMC (%)	7.7 ± 2.7	7.6 ± 2.5	7.4 ± 4.4	8.4 ± 5	6.4 ± 3.8	7.5 ± 3.6	6.7 ± 4.1	4 ± 4.4
CD3-CD56 + NK cells of PBMC (%)	10 ± 4.6	10 ± 4.7	8.4 ± 5.1	9.2 ± 5	7.8 ± 4.2	8.7 ± 4.7	8.6 ± 5.2	8.6 ± 5
CD14 + Monocytes of PBMC (%)	17 ± 4.5	18 ± 6.1	18 ± 8.5	20 ± 8.3	18 ± 11	19 ± 10	19 ± 8.3	20 ± 7.9
Has RNA-seq	97% (59)	52% (32)	74% (68)	65% (60)	75% (44)	80% (47)	74% (25)	74% (25)
Has genotyping	93% (57)	93% (57)	95% (87)	95% (87)	93% (55)	93% (55)	91% (31)	91% (31)

CCP: cyclic citrullinated peptide, CRP: C-reactive protein, RF: rheumatic factor, DAS28-CRP: disease activity score 28, ESR: erythrocyte sedimentation rate, PBMC: peripheral blood mononuclear cells, RF: rheumatic factor, VAS: visual analog scale.

Numerical variables are reported as mean ± SD and categorical variables are reported as percentage (count).

Cohort HC: healthy controls.

Cohort A: treatment naïve, initiating Methotrexate.

Cohort B: methotrexate non-responders, initiating TNFi.

Cohort C: TNFi non-responders, initiating second biological (Supplementary Figure S2).

not based on the arbitrary median cutoffs did not provide any ability to stratify patient groups. The number of Tocilizumab treated patients in the cohort was too low to allow separate replication of the Tocilizumab findings in Dennis *et al.* Using only the levels of these two proteins to predict anti-TNF-response of  $\Delta$ DAS28-CRP below  $-1.2$ , we obtained a Receiver Operating Characteristic (ROC)-curve with an Area Under Curve (AUC) of 0.714.

### Genetic Associations with Anti-TNF-Response

We identified three published studies that identified SNPs associated with anti-TNF-response (11–13), as well as the results from the winning algorithm at the recent Rheumatoid Arthritis Responder Challenge, developed by team GuanLab, University of Michigan (<https://www.synapse.org/#!Synapse:syn2504648>). Altogether 52 potentially anti-TNF-response associated

SNPs were considered. Since the COMBINE database is more than 10 times smaller than current cohorts for genetic association studies, it is underpowered for a formal validation of these SNPs. Nonetheless we tested all 52 SNPs for association to  $\Delta$ DAS28-CRP after treatment with TNF-blocking agents for 3 months. Only two SNPs were found to be associated with  $\Delta$ DAS28-CRP in COMBINE, one rs6028945 (13) ( $P = 0.047$ ) and one rs7305646 (12) ( $P = 0.014$ ).

### Integration of Known Biomarkers for Anti-TNF-Response Prediction

To provide an integrative approach for prediction of anti-TNF treatment, we tested the prediction level obtainable when combining the best current high-throughput methodology and published findings replicated in the COMBINE cohort. In the test we include the expression of the 8 genes from transcriptomics literature, the sICAM1/CXCL13 finding (14) and the two SNPs

from GWAS literature (12,13). Combining these 11 variables gives a model that explains 51% of the variation in  $\Delta$ DAS28-CRP in the TNFi group of the COMBINE cohort (linear regression, adjusted  $R^2$ ). The largest explanatory power came from the sICAM1/CXCL13 finding as well as the two most significant measures of gene expression levels (CX3CR1 and SLC2A3). If the  $\Delta$ DAS28-CRP is converted to a binary response with  $-1.2$  as cutoff, a ROC-curve based on the fitted values yields an AUC of 0.815 (Figure 2D). An overview of all response and model definitions is given in the Supplement 3. This includes a comparison with age–gender–smoking status as well as ACR good response definitions.

### DISCUSSION

The main finding in our study was in replication of 11 biomarkers for anti-TNF response in RA and in successful integration of these biomarkers in a single model with high predictive value.

We collected double-time point blood samples from 185 RA patients, and acquired clinical information, transcriptomics (RNA-sequencing), genomics (genotyping microarrays) and proteomics (HumanMAP and VectraDA) data. Each of these steps was described in detail herein and an extensive quality control was performed. We chose to strictly investigate already published biomarkers for anti-TNF treatment response as a first stage to prove validity and usefulness of COMBINE biobank. To ultimately benefit patients, it is essential that such validation studies are performed (20). Additionally, it is noteworthy that a voluntary restriction of the search space for our investigations limits the otherwise large problems with multiple testing burdens and underpowered studies that are typically seen in high-throughput discovery studies.

The serum protein level of two previously proposed biomarkers had the strongest impact on prediction (sICAM1 and CXCL13). This was closely followed in strength by the PBMC gene expression of two previously suggested transcriptomic biomarkers. The knowledge gained from genetic markers, in this case two previously proposed SNPs, only improved predictions in a comparatively lesser fashion. Keeping in mind that the COMBINE study design was fundamentally method agnostic, this has an important impact on where to focus future efforts in biomarker discovery.

Specifically for the sICAM1 and CXCL13 serum protein levels it is of interest that these have already been extensively discussed as biomarkers in the literature: high baseline plasma CXCL13 levels in RA patients are associated with an improved chance of remission after 2 years (21), and it has also been proposed that high concentrations of CXCL13 in plasma indicate recent onset of inflammation that may respond better to early aggressive treatment. On the other hand, high expression levels of CXCL13 gene in synovium of RA

patients correlates with more severe disease (22). Likewise, serum CXCL13 correlates with synovial CXCL13 measured at a single joint, suggesting synovitis as an important source of circulating CXCL13. Cell and animal models suggest the implication of ICAM1 in RA (23), but comparable evidence of levels in serum and synovial tissue of ICAM1 in RA patients is not available.

The set of 11 SNPs, genes and proteins that were shown to be predictive in this study represents a step forward toward use of anti-TNF-response-predictive signatures in a clinical setting. In addition to being individually replicated, their compound prediction accuracy exceeds what was previously observed. While the prediction accuracy is still below that of precision medicine in cancer (for example, Herceptin), it is still conceivable that a sizeable fraction of anti-TNF-initiating patients could be selected who would benefit from progressing directly to other biologic treatments now available. To accurately assess the size of this, however, a prospective study with a pre-defined biomarker signature is required.

## CONCLUSION

To our knowledge this is the first time a combined multi-omics approach has been taken to the question of drug-response stratification in RA. The prediction strength was calculated as an AUC of 0.815 that is relatively high for pharmacogenomics studies.

In conclusion we present these results as a comment to the state of precision medicine, providing evidence that a few blood biomarkers may serve as a prognostic tool for the prediction of clinical response of TNF blocking therapy. If a prediction proficiency of this magnitude is introduced in a clinical setting, it may have a strong impact on established clinical treatment routines.

## ACKNOWLEDGMENTS

We wish to thank all the participating patients and healthy controls and the research nurse Karin Lundvall.

## DISCLOSURE

The vast majority of expenses related to this study including laboratory measurements and study costs were funded by the pharmaceutical company Novo Nordisk A/S, at which the authors LF and UGWM were employed at the time of the data collection. However, the authors wish to make clear that this company had no influence on any conclusions in this study, nor are any authors currently affiliated or employed by this company. The stratification potential of this prediction signature is being investigated for patenting as US application number 14947077 and UK application number 1520524.8.

## REFERENCES

1. Cronstein BN (2005) Low-dose methotrexate: a mainstay in the treatment of rheumatoid arthritis. *Pharmacol. Rev.* 57(2):163–72.
2. Feldmann M, Maini RN (2001) Anti-TNF alpha therapy of rheumatoid arthritis: what have we learned? *Annu Rev Immunol* 19:163–96.
3. Koczan D, Drynda S, Hecker M, Drynda A, Guthke R, Kekow J, et al. (2008) Molecular discrimination of responders and nonresponders to anti-TNF alpha therapy in rheumatoid arthritis by etanercept. *Arthritis Res Ther* 10(3):R50.
4. Lequerre T, Gauthier-Jauneau AC, Bansard C, Derambure C, Hiron M, Vittecoq O, et al. (2006) Gene profiling in white blood cells predicts infliximab responsiveness in rheumatoid arthritis. *Arthritis Res Ther* 8(4):R105.
5. Sekiguchi N, Kawauchi S, Furuya T, Inaba N, Matsuda K, Ando S, et al. (2008) Messenger ribonucleic acid expression profile in peripheral blood cells from RA patients following treatment with an anti-TNF-alpha monoclonal antibody, infliximab. *Rheumatology (Oxford)* 47(6):780–8.
6. Julia A, Erra A, Palacio C, Tomas C, Sans X, Barcelo P, et al. (2009) An eight-gene blood expression profile predicts the response to infliximab in rheumatoid arthritis. *PLoS One* 4(10):e7556.
7. Tanino M, Matoba R, Nakamura S, Kameda H, Amano K, Okayama T, et al. (2009) Prediction of efficacy of anti-TNF biologic agent, infliximab, for rheumatoid arthritis patients using a comprehensive transcriptome analysis of white blood cells. *Biochem. Biophys. Res. Commun.* 387(2):261–5.
8. Stuhlmüller B, Haupl T, Hernandez MM, Grutzkau A, Kuban RJ, Tandon N, et al. (2010) CD11c as a transcriptional biomarker to predict response to anti-TNF monotherapy with adalimumab in patients with rheumatoid arthritis. *Clin Pharmacol. Ther.* 87(3):311–21.
9. Oliveira RD, Fontana V, Junta CM, Marques MM, Macedo C, Rassi DM, et al. (2012) Differential

- gene expression profiles may differentiate responder and nonresponder patients with rheumatoid arthritis for methotrexate (MTX) monotherapy and MTX plus tumor necrosis factor inhibitor combined therapy. *J Rheumatol* 39(8):1524–32.
10. Toonen EJ, Gilissen C, Franke B, Kievit W, Eijsbouts AM, den Broeder AA, *et al.* (2012) Validation study of existing gene expression signatures for anti-TNF treatment in patients with rheumatoid arthritis. *PLoS One* 7(3):e33199.
  11. Cui J, Stahl EA, Saevardottir S, Miceli C, Diogo D, Trynka G, *et al.* (2013) Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis. *PLoS Genet.* 9(3):e1003394.
  12. Plant D, Bowes J, Potter C, Hyrich KL, Morgan AW, Wilson AG, *et al.* (2011) Genome-wide association study of genetic predictors of anti-tumor necrosis factor treatment efficacy in rheumatoid arthritis identifies associations with polymorphisms at seven loci. *Arthritis Rheum.* 63(3):645–53.
  13. Liu C, Batliwalla F, Li W, Lee A, Roubenoff R, Beckman E, *et al.* (2008) Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Mol. Med.* 14(9–10):575–81.
  14. Dennis G, Jr., Holweg CT, Kummerfeld SK, Choy DF, Setiadi AF, Hackney JA, *et al.* (2014) Synovial phenotypes in rheumatoid arthritis correlate with response to biologic therapeutics. *Arthritis Res Ther.* 16(2):R90.
  15. Soderlin MK, Petersson IF, Geborek P (2012) The effect of smoking on response and drug survival in rheumatoid arthritis patients treated with their first anti-TNF drug. *Scand J Rheumato.* 141(1):1–9.
  16. Kim D, Perteu G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14(4):R36.
  17. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26(1):139–40.
  18. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, *et al.* (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform.* 14(6):671–83.
  19. Gadin JR, van't Hooft FM, Eriksson P, Folkersen L (2015) AllelicImbalance: an R/bioconductor package for detecting, managing, and visualizing allele expression imbalance data from RNA sequencing. *BMC Bioinformatics.* 16:194.
  20. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, *et al.* (2014) Increasing value and reducing waste in research design, conduct, and analysis. *Lancet.* 383(9912):166–75.
  21. Greisen SR, Schelde KK, Rasmussen TK, Kragstrup TW, Stengaard-Pedersen K, Hetland ML, *et al.* (2014) CXCL13 predicts disease activity in early rheumatoid arthritis and could be an indicator of the therapeutic ‘window of opportunity’. *Arthritis Res Ther.* 16(5):434.
  22. Bugatti S, Manzo A, Vitolo B, Benaglio F, Binda E, Scarabelli M, *et al.* (2014) High expression levels of the B cell chemoattractant CXCL13 in rheumatoid synovium are a marker of severe disease. *Rheumatology (Oxford)* 53(10):1886–95.
  23. Ahmed S, Riegsecker S, Beamer M, Rahman A, Bellini JV, Bhansali P, *et al.* (2013) Largazole, a class I histone deacetylase inhibitor, enhances TNF-alpha-induced ICAM-1 and VCAM-1 expression in rheumatoid arthritis synovial fibroblasts. *Toxicol Appl Pharmacol.* 270(2):87–96.

Cite this article as: Folkersen L, *et al.* (2016) Integration of known DNA, RNA and protein biomarkers provides prediction of anti-TNF response in rheumatoid arthritis: Results from the COMBINE study. *Mol. Med.* 22:322–8.